Document Metadata Review

What's the cheapest way to achieve the following common tasks: finding, determining the latest revision, and listing the works of an author?

Digital Object Identifiers

Digital Object Identifiers (DOI) seem popular, and while it's a federated Handle system, the federation charges upkeep fees so that the cheapest document registrar charges 1 USD per document + 275 USD per year.[1]

Citation Managers

webcite[2] has become for-profit, and neither perma.cc[3] nor archive.is[4] handles PDFs. And none afford determining document revision status.

The Internet Archive

PURL[5] has been adopted by the Internet Archive[6] (IA), however one would have to periodically manage the HTTP 301 redirects as needed.

IA also provides Archival Resource Keys (ARKs) which are similar to DOIs in that one can request the metadata via the ARK Description service, and one can also pull the original resource via the Access service.

urn:uuid:...

As version 4 UUIDs allow for decentralized issuance, generate one UUID to identify the author, and subsequently one UUID for each paper.

Create a PDF/A with the UUID metadata, and submit it to archive.org or webcite. It will then be indexed by google.

Caveats to this include: how does a reader determine which version is latest?

ARK + UUID

We can self-publish our documents to a high-volume primary distribution facility, and then to IA for permanent storage.

IA allows users to set the ARK identifier on upload[7] which allows the use of locally generated UUIDs as ARKs. The workflow then becomes: create the document with PDF Info metadata and UUID assigned, then upload it to IA. Any time there is a change, publish the new document to IA using the same ARK, and IA will maintain a history of the revisions.

---

[1] https://www.crossref.org/fees/

[2] http://webcitation.org

[3] http://perma.cc

[4] http://archive.is

[5] http://purl.org

[6] http://archive.org

[7] https://blog.archive.org/2013/02/08/presetting-metadata-with-the-new-beta-uploader/